well as temporal redundancy, should be required. Another interesting analysis perspective can be to directly analyze the $3D+t$ data as a 4-D volume both for SHG and THG. In summary, *in vivo* SHG and THG microscopy represents a new powerful bioimaging modality and new processing methodologies have to be developed in consonancy [18].

VII. Conclusion

In this paper, a dedicated image acquisition and processing pipeline for tracking and segmentation of SHG/THG image sequences of early zebrafish development has been designed and implemented. This methodology has been used for obtaining the digital reconstruction of six embryos including complete division timings, cell coordinates, and cell shape until the 1000-cell stage with minute temporal accuracy and micrometer spatial resolution. The data analysis of the reconstructed embryos provide a new comprehensive quantitative description of early zebrafish development [8].

References

[1] A. Oates, N. Gorfinkiel, M. González-Gaitán, and C.-P. Heisenberg, "Quantitative approaches in developmental biology," *Nat. Rev. Genet.*, vol. 10, no. 8, pp. 517–530, Aug. 2009.

[2] C. Zanella, M. Campana, B. Rizzi, C. Melani, G. Sanguinetti, P. Bourgine, K. Mikula, N. Peyrieras, and A. Sarti, "Cells segmentation from 3-D confocal images of early zebrafish embryogenesis," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 770–781, Mar. 2010.

[3] M. Luengo-Oroz, M. Ledesma-Carbayo, N. Peyriéras, and A. Santos, "Image analysis for understanding embryo development: A bridge from microscopy to biological insights," *Curr. Opin. Genet. Dev.*, vol. 21, no. 5, pp. 630–637, Oct. 2011.

[4] J. Amatruda, J. L. Shepard, H. M. Stern, and L. I. Zon, "Zebrafish as a cancer model system," *Cancer Cell*, vol. 1, no. 3, pp. 229–231, Apr. 2002.

[5] P. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. K. Stelzer, "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy," *Science*, vol. 322, no. 5904, pp. 1065–1069, Nov. 2008.

[6] D. Kane and C. B. Kimmel, "The zebrafish midblastula transition," *Development*, vol. 119, no. 2, pp. 447–456, Oct. 1993.

[7] C. Sun, S. W. Chu, S. Y. Chen, T. H. Tsai, T. M. Liu, C. Y. Lin, and H. J. Tsai, "Higher harmonic generation microscopy for developmental biology," *J. Struct. Biol.*, vol. 147, no. 1, pp. 19–30, Jul. 2004.

[8] N. Olivier, M. A. Luengo-Oroz, L. Duloquin, E. Faure, T. Savy, I. Veilleux, X. Solinas, D. Débarre, P. Bourgine, A. Santos, N. Peyriéras, and E. Beaurepaire, "Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy," *Science*, vol. 329, no. 5994, pp. 967–971, Aug. 2010.

[9] M. Luengo-Oroz, T. Savy, J. L. Rubio, L. Duloquin, E. Faure, N. Olivier, M. Ledesma-Carbayo, D. Debarre, P. Bourgine, E. Beaurepaire, N. Peyrieras, and A. Santos, "Processing pipeline for digitalizing the lineage tree of early zebrafish embryogenesis from multiharmonic imaging," in *Proc. 8th IEEE Int. Symp. Biomed. Imag.*, 2011, pp. 1561–1564.

[10] D. Débarre, W. Supatto, E. Farge, B. Moulia, M. C. Schanne-Klein, and E. Beaurepaire, "Velocimetric third-harmonic generation microscopy: Micrometer-scale quantification of morphogenetic movements in unstained embryos," *Opt. lett.*, vol. 29, no. 24, pp. 2881–2883, Dec. 2004.

[11] D. Débarre, W. Supatto, A. M. Pena, A. Fabre, T. Tordjmann, L. Combettes, M. C. Schanne-Klein, and E. Beaurepaire, "Imaging lipid bodies in cells and tissues using third-harmonic generation microscopy," *Nat. Methods*, vol. 3, no. 1, pp. 47–53, Jan. 2006.

[12] W. Zipfel, R. M. Williams, R. Christie, A. Y. Nikitin, B. T. Hyman, and W. W. Webb, "Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation," *Proc. Nat. Acad. Sci.*, vol. 100, no. 12, pp. 7075–7080, Jun. 2003.

[13] M. Luengo-Oroz, L. Duloquin, C. Castro, T. Savy, E. Faure, B. Lombardot, R. Bourgine, N. Peyrieras, and A. Santos, "Can Voronoi diagram model cell geometries in early sea-urchin embryogenesis?," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 504–507.

[14] C. Vachier and F. Meyer, "The viscous watershed transform," *Math. Imag. Vis.*, vol. 22, no. 2/3, pp. 251–267, May 2005.

[15] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3-D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[16] D. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[17] I. Smal, M. Loog, W. Niessen, and E. Meijering, "Quantitative comparison of spot detection methods in fluorescence microscopy," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 282–301, Feb. 2010.

[18] D. Evanko, "Microscope harmonies," *Nat. Methods*, vol. 7, no. 10, p. 779, Oct. 2010.

Particle Filter With a Mode Tracker for Visual Tracking Across Illumination Changes

Samarjit Das, Amit Kale, and Namrata Vaswani

***Abstract*—In this correspondence, our goal is to develop a visual tracking algorithm that is able to track moving objects in the presence of illumination variations in the scene and that is robust to occlusions. We treat the illumination and motion (x-y translation and scale) parameters as the unknown "state" sequence. The observation is the entire image, and the observation model allows for occasional occlusions (modeled as outliers). The nonlinearity and multimodality of the observation model necessitate the use of a particle filter (PF). Due to the inclusion of illumination parameters, the state dimension increases, thus making regular PFs impractically expensive. We show that the recently proposed approach using a PF with a mode tracker can be used here since, even in most occlusion cases, the posterior of illumination conditioned on motion and the previous state is unimodal and quite narrow. The key idea is to importance sample on the motion states while approximating importance sampling by posterior mode tracking for estimating illumination. Experiments demonstrate the advantage of the proposed algorithm over existing PF-based approaches for various face and vehicle tracking. We are also able to detect illumination model changes, e.g., those due to transition from shadow to sunlight or vice versa by using the generalized expected log-likelihood statistics and successfully compensate for it without ever loosing track.**

***Index Terms*—Monte Carlo methods, particle filter (PF), tracking, visual tracking.**

I. Introduction

In recent works [3]–[5], we developed practically implementable particle filtering (PF) approaches for tracking on large-dimensional state spaces with multimodal likelihoods. An approach called PF with posterior mode tracking (MT) was introduced. The focus of [3] was only on the algorithm design and analysis, and we only showed one simulated temperature field tracking problem. In this correspondence,

Manuscript received December 06, 2010; revised July 12, 2011; accepted October 13, 2011. Date of publication November 03, 2011; date of current version March 21, 2012. This work was presented in part at the 2007 International Conference on Acoustics, Speech, and Signal Processing; and at the 2008 International Conference on Image Processing. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wan-Chi Siu.

S. Das and N. Vaswani are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: samarjit@iastate.edu; namrata@iastate.edu).

A. Kale is with Medical Imaging Technologies team, Siemens Corporate Technology, Bangalore 560100, India (e-mail: kale.amit@siemens.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2174370

we look at the problem of moving objects' tracking across illumination change, which is a key practical application where the above problem occurs. We explain how to use the particle filter with a mode tracker (PF-MT) approach to design an efficient PF-based tracker for this problem. A significantly improved performance of PF-MT over existing PF approaches as well as over some other illumination tracking approaches is demonstrated. We note that this is the first work where the PF-MT approach is exhaustively compared against existing PF methods for a real visual tracking application. The only other work where PF-MT has been used for a real application is in [6] but that only compared two different contour deformation models and not PF algorithms. We also briefly demonstrate the use of the expected log-likelihood (ELL) statistics [7] to detect the illumination model change and to automatically adapt to that change. This is needed, e.g., when the moving target (vehicle or persons) moves from a lighted to a shadowed region or vice versa.

In the absence of illumination change, the motion of a rigid object moving in front of a camera that is far from the scene can be tracked using a 3-D vector consisting of x-y translation and uniform scale or, more generally, using a 6-D affine model as in [8]. *Condensation* was the first work to beautifully demonstrate the use of a PF for tracking through multimodal observation likelihoods resulting from background clutter or occlusions. Now, if illumination also changes over time and if different parts of the object experience different lighting conditions, then many more dimensions get added to the state space, thus making it a much larger dimensional problem. As the state-space dimension increases, the number of particles required to track using a PF increases [9], thus making PF impractical. However, as we explain later in Section III, in most cases of practical interest, the posterior of illumination change, which is conditioned on motion and on the previous state, is unimodal. In addition, illumination change is usually very gradual, and this causes the posterior to be also narrow. Under these two assumptions, one can replace importance sampling of illumination by just posterior MT, i.e., we can use a PF-MT for this problem. This one step reduces the importance sampling dimension to three and thus drastically reduces the number of particles required. We refer to the resulting PF as the illumination PF-MT.

Early works on illumination modeling for object recognition and illumination-invariant tracking include [10]–[13]. However, learning these models from low-resolution videos might pose serious practical challenges. Template adaptation approaches [14]–[16] suffer from problems of drift [17], e.g., if you adapt when the tracker has latched onto clutter, it will lead to tracking failure. In [18], it was assumed that a small number of centroids in the illumination space can be used to represent the illumination conditions existing in the scene. The six-centroid method of [18] does not suffice given complex illumination patterns often encountered in reality. In addition, it is unclear how standard trackers such as a mean-shift tracker [19] can be adapted for illumination invariance. We show examples of failure of template adaptation, mean shift, and the six-centroid method in Fig. 3. Some recent work on jointly handling appearance change due to illumination variations, as well as other factors such as a 3-D pose change, include [20]–[25].

In this note, we use a template-based tracking framework because it is simple to use and to explain our key ideas; but the illumination PF-MT approach can also be used with other representations of the target, e.g., feature-based approaches. In addition, the template matching framework enables illumination to be parameterized using a Legendre basis, as suggested in past works [18], [26]. We use a very simple motion and illumination change model to demonstrate how to design the PF-MT for our problem. However, we note that a similar PF-MT idea can also be extended to jointly handle appearance change due to illumination as well as other factors such as 3-D pose change by using the more sophisticated models of recent work [20]–[23], [25]. Similarly, illumination can also be represented using other parameterizations such as those proposed in [10]–[13].

In Section II, we develop the state-space model for the tracker. In Section III, we explain how to design a PF-MT algorithm for tracking across illumination variations. In Section IV, we design the generalized ELL-based (gELL) change detection system for detecting and tracking illumination model changes. The experimental results are given in Section V, and we conclude this correspondence in Section VI.

II. State-Space Model

The system model consists of simple dynamical models for illumination Λ_t and motion parameters U_t (further details to follow). Thus, the state vector is given as $X_t = [U_t^T, \Lambda_t^T]^T$. Observation Y_t is the image frame at time t. Our observation model assumes that occlusions may occur.

A. Notation

Notation $\mathrm{vec}(.)$ refers to the vectorization operator, which operates on a $m \times n$ matrix to give the vector of dimension mn by cascading the rows. $[x]_n$ denotes the nth element of vector x. $\mathbf{I}$ denotes the identity matrix. The Hadamard product (the ".*" operation in MATLAB) is denoted by $\odot$. The terms $\mathbf{1}$ and $\mathbf{0}$ refer to the column vectors with all entries as 1 and 0, respectively. mean(.) gives the mean value of the entries of a vector, i.e., $\mathrm{mean}(x) = 1/N \sum_i^N [x]_i$ for an N-length vector x. Function round($\mathbf{Z}$) operates on matrix $\mathbf{Z}$ and outputs a matrix with integer entries as round($z_{i,j}$), which is the integer closest to $z_{i,j}$. Notation $\mathcal{N}(a; \mu, \Sigma)$ denotes the value of the Gaussian distribution with mean μ and covariance Σ computed at a, whereas $x \sim \mathcal{N}(\mu, \Sigma)$ implies that the random variable x is Gaussian distributed with mean μ and covariance Σ. Similarly, notation $\mathcal{U}(a; c_1, c_2)$ denotes the value of the uniform density defined over $[c_1, c_2]$ computed at a, whereas $x \sim \mathcal{U}(c_1, c_2)$ denotes that x is uniformly distributed over $[c_1, c_2]$. The term mode refers to the local maxima of a probability density function (pdf). The pdf is unimodal (multimodal) if it has exactly one (multiple) mode(s).

B. Observation Model

We assume the observation model of [1] and [18] but include an occlusion model similar to the one introduced in condensation [8]. The target object template image, at time t, is denoted as T_t. As introduced in [18], the changed "appearance" of template T_t is represented in terms of a linear combination of the initial template T_0 scaled by a set of Legendre basis functions as follows:

$$\mathrm{vec}(T_t) = A\Lambda_t \tag{1}$$

where matrix A is defined in (7). Its columns consist of the initial template scaled by the first D Legendre basis functions. The $D \times 1$ vector Λ_t is the Legendre basis coefficients at time t along the first D Legendre basis functions. Henceforth, we will call it the illumination vector. Theoretically, illumination can be different for each pixel, and the illumination dimension would become equal to M. However, it is known from earlier work that, in reality, the variability is not so high and that the top Legendre coefficients suffice to model most of the illumination changes [18]. The Legendre basis coefficients were successfully used in [1]. In this correspondence, we use the Legendre basis, although any other suitable basis (even data-dependent basis such as PCA) can be used as well, and nothing in our proposed algorithm will change. It is to be noted that we only model translation and scaling of the template as the motion states. Note that the template is not updated as a whole but we update the motion states that need to be applied to

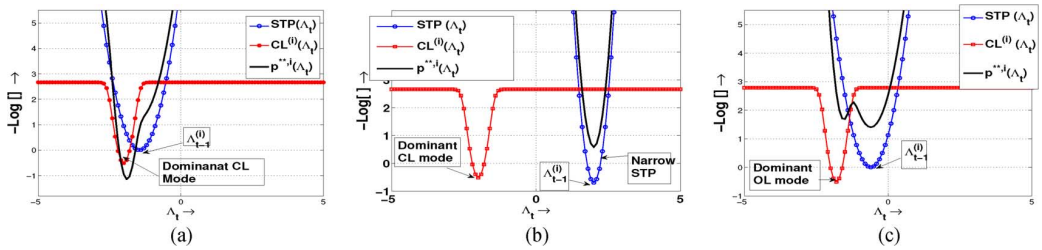

Fig. 1. In this figure, we plot the negative log of the CL, i.e., $\text{CL}^{(i)}(\Lambda_t)$ and STP $p(\Lambda_t|\Lambda_{t-1}^{(i)})$ together with the negative log of $p^{**,i}$ for a simple 1-D scalar case. (a) No occlusion case where the CL mode is very close to the STP mode leading to unimodal $p^{**,i}$. (b) Occlusion case with unimodal $p^{**,i}$ (the CL mode is very far from the STP mode). (c) CL mode is close to the STP mode (but not in the basin of attraction of the STP); this gives rise to a multimodal $p^{**,i}$. (a) Unimodal $p^{**,i}(\Lambda_t)$. (b) Unimodal $p^{**,i}(\Lambda_t)$. (c) Multimodal $p^{**,i}(\Lambda_t)$.

the original template to get something that matches the object in the current image.

Given the motion parameter vector U_t consisting of scale, horizontal translation, and vertical translation ($U_t = [s_t\ \tau_t^h\ \tau_t^v]^T$) of the initial template, the "moved"(translated/scaled) template region in the current frame Y_t can be computed, as given in (8) in the Appendix. We call it the region of interest or ROI(U_t).

At any given instant, part or all of the ROI may get covered (occluded) by some other object(s). In the absence of any knowledge about the occluding object(s)'s intensity or pixel locations, we assume a simple outlier noise model for occlusion [8]. At any time t, any ROI pixel gets occluded with probability $(1-\theta)$ independent of the others and when it does, its intensity is uniformly distributed between 0 and 255, independent of all other pixels. On the other hand, with probability θ, there is no occlusion. Thus, for all $i \in \text{ROI}(U_t)$

$$Y_t(i) = \begin{cases} [A\Lambda_t]_i + [\mathbf{n_t}]_i & w.p \quad \theta \\ [O_t]_i & w.p \quad 1-\theta \end{cases}$$

where $\mathbf{n_t} \sim \mathcal{N}(0, \sigma_o^2\mathbf{I})$, $[O_t]_i \sim \mathcal{U}(0,255)$, and ROI($U_t$) is computed using (8). The pixels outside the predicted template region (ROI) are assumed to have intensities that do not depend on U_t, Λ_t, or $T_0(i,j)$. Thus, we have the following observation likelihood given the state vector $X_t \triangleq [U_t^T, \Lambda_t^T]^T$:

$$\begin{aligned} \text{OL}(X_t) &\triangleq p(Y_t|U_t,\Lambda_t) \propto p\left(Y_t\left(\text{ROI}(U_t)\right)|\Lambda_t\right) \\ &= \Pi_{n=1}^{M}\left[\theta\,\mathcal{N}\left([Y_t\left(\text{ROI}(U_t)\right)]_n;(A\Lambda_t)_n,\sigma_o^2\right)\right. \\ &\qquad \left. + (1-\theta)\mathcal{U}\left([Y_t\left(\text{ROI}(U_t)\right)]_n;0,255\right)\right] \end{aligned} \tag{2}$$

where $[\,]_n$ denotes the nth element of a vector. Note that the outlier noise term in (2) does not depend on X_t; thus, each of the M terms in the product is a heavy-tailed probability distribution function and, hence, multimodal. For the given realization $U_t^{(i)}$ of U_t, we define the conditional likelihood (CL) of Λ_t as

$$\text{CL}^{(i)}(\Lambda_t) \triangleq \text{OL}\left(\Lambda_t, U_t^{(i)}\right). \tag{3}$$

An example of the negative-log plot of $\text{CL}^{(i)}$ for a scalar case (i.e., $M=1$) is shown in Fig. 1.

C. System Model

We defined the motion parameter vector $U_t = [s_t\ \tau_t^x\ \tau_t^y]^T$ in the previous section. As mentioned earlier, the illumination vector $\Lambda_t \in \mathcal{R}^D$ corresponds to the coefficients of the Legndre basis function. Thus, tracking is performed over a $D+3$-dimensional motion–illumination space.

In the absence of specific information about the object motion or about illumination variation, we assume a simple random-walk model on both U_t and Λ_t, i.e.,

$$U_t = U_{t-1} + n_u \quad \text{and} \quad \Lambda_t = \Lambda_{t-1} + n_\lambda \tag{4}$$

where $n_u \sim \mathcal{N}(0,\Sigma_u)$, $n_\lambda \sim \mathcal{N}(0,\Sigma_\Lambda)$, Σ_Λ is a $D \times D$ diagonal matrix, and Σ_u is a 3×3 diagonal covariance matrix. Thus, the state transition prior (STP) can be given as

$$\begin{aligned} &\text{STP}(U_t,\Lambda_t;U_{t-1},\Lambda_{t-1}) \triangleq \text{STP}(U_t;U_{t-1})\ STP(\Lambda_t;\Lambda_{t-1}) \\ &\text{where} \quad \text{STP}(U_t;U_{t-1}) \triangleq \mathcal{N}(U_t;U_{t-1},\Sigma_u) \\ &\text{and} \quad \text{STP}(\Lambda_t;\Lambda_{t-1}) \triangleq \mathcal{N}(\Lambda_t;\Lambda_{t-1},\Sigma_\Lambda). \end{aligned} \tag{5}$$

III. Illumination PF-MT

A PF uses sequential importance sampling [9] along with a resampling step [27] to empirically estimate the posterior distribution $\pi_{t|t}(X_t) \triangleq p(X_t|Y_{1:t})$ of state X_t. PF-MT [3] splits the state vector X_t into $X_t = [X_{t,s}, X_{t,r}]$, where $X_{t,s}$ denotes the coefficients of a small-dimensional "effective basis" (in which most of the state change is assumed to occur), whereas $X_{t,r}$ belongs to the "residual space" in which the state change is assumed to be "small." It importance samples only on the effective basis dimensions but replace importance sampling by deterministic posterior MT in the residual space. Thus, the importance sampling dimension is only $\dim(X_{t,s})$ (much smaller than $\dim(X_t)$), and this is what decides the effective particle size. PF-MT implicitly assumes that the posterior of the residual space conditioned on the effective basis ("conditional posterior") is unimodal most of the time. Moreover, it is also assumed to be narrow. Only under these two assumptions that the conditional posterior mode is a highly likely sample from the conditional posterior.

Consider the observation model given in (2). Without any prior information, due to clutter, the observation likelihood $\text{OL}(U_t,\Lambda_t)$ is clearly multimodal, e.g., if there is no constraint on how large illumination change can be, one may get a very strong match to the observation with a wrong object region (wrong motion estimate). This necessitates the use of a PF. In addition, even with just a 7-D space of illumination change, the state-space dimension becomes ten, which is quite large. As a result, the original PF [27] will require a very large number of particles. Other efficient PFs such as PF-Doucet [9] or Gaussian PF [28] cannot also be used since these implicitly assume that the posterior conditioned on the previous state $p^*(X_t) \triangleq p(X_t|X_{t-1},Y_t)$ is unimodal. However, in our problem, this will not hold since the likelihood is multimodal and since the prior on motion is typically quite broad, and this will result in a multimodal p^* (as explained in [3]). In fact, PF-Doucet [9] cannot even be implemented easily because it requires finding the posterior

mode (mode of OL(X_t)STP(X_t)). However, since our OL(X_t) is not differentiable [consists of a round operation, see (8)], one cannot use standard numerical optimization algorithms to do this. Moreover, since the multimodality in the problem comes from the likelihood (and not the system model), at any time, Gaussian mixture filters or Gaussian Sum PF [29] also cannot be used (see discussion in [3] for details). Because of the occlusion term, even conditioned on motion U_t, the observation model is not linear Gaussian. Hence, Rao–Blackwellized PF (RB-PF [30], [31]) cannot be used either.

However, notice that, while p^* is often multimodal, p^* is conditioned on motion, i.e.,

$$p^{**,i}(\Lambda_t) \triangleq p^*\left(X_t|U_t^{(i)}\right) = p\left(\Lambda_t|X_{t-1}^{(i)}, Y_t, U_t^{(i)}\right) \propto \mathrm{CL}^{(i)}(\Lambda_t)p\left(\Lambda_t|\Lambda_{t-1}^{(i)}\right) \quad (6)$$

is usually unimodal. Here, $\mathrm{CL}^{(i)}(\Lambda_t)$, defined in (3), is the CL of Λ_t, i.e., $p(Y_t|U_t^{(i)}, \Lambda_t)$. This happens for the following reason. Notice that $p(\Lambda_t|\Lambda_{t-1}^{(i)})$ is Gaussian and, hence, unimodal. When there is no occlusion, the dominant CL mode is the observed illumination of the target. Hence, the STP's mode (target illumination at $t-1$) is close to it and in fact lies in its basin of attraction; therefore, $p^{**,i}(\Lambda_t)$ is unimodal [see Fig. 1(a)]. In case of occlusion, the dominant CL mode is the intensity pattern of the occluding object. However, since the illumination change prior is quite narrow, the conditional posterior $p^{**,i}$ will still be unimodal [see Fig. 1(b)], except if the occlusion intensity is very close to the target's intensity pattern [see Fig. 1(c)]. This fact is proved in [6, Theorem 1]. In both occlusion and no-occlusion cases, the narrowness of STP(Λ_t) ensures narrowness of $p^{**,i}$. As a result, we can use the PF-MT for this problem with $X_{t,s} = U_t$ and $X_{t,r} = \Lambda_t$. We give the stepwise illumination PF-MT algorithm in Algorithm 1. The only exception where the above split up may not work is if the occluding object's intensity pattern is very close to that of the template i.e., the case of Fig. 1(c). If in an application, this happens very often, then one should also use a part of the illumination state as $X_{t,s}$.

Algorithm 1 Illumination PF-MT. Going from $\pi_{t-1|t-1}^N$ to $\pi_{t|t}^N(X_t) = \sum_{i=1}^N w_t^{(i)}\delta(X_t - X_t^{(i)})$

For each $t > 0$,

1) *Importance Sample on motion*: For all i, sample $U_t^{(i)} \sim \mathrm{STP}(U_t; U_{t-1}^{(i)})$ [defined in (5)]. Use $U_t^{(i)}$ to compute the corresponding ROI using (8).
2) *MT on illumination*: Use the current observation to get $Y_t(\mathrm{ROI}(U_t^{(i)}))$ and compute the mode $m_t^{(i)}$ of $p^{**,i}(\Lambda_t)$ by solving the following convex optimizing problem:

$$m_t^{(i)} = \arg\min_{\Lambda_t}\left[-\log p^{**,i}(\Lambda_t)\right] = \arg\min_{\Lambda_t} L^{(i)}(\Lambda_t)$$

$$\text{where} \quad L^{(i)}(\Lambda_t) = \left[-\log \mathrm{CL}^{(i)}(\Lambda_t)\right] + \left[-\log \mathrm{STP}\left(\Lambda_t; \Lambda_{t-1}^{(i)}\right)\right]$$

where $\mathrm{CL}^{(i)}(\Lambda_t)$ is defined in (3) and STP(.) in (5). Generate the illumination particle as $\Lambda_t^{(i)} = m_t^{(i)}$
3) *Weighting and Resampling*: Compute the weights using $w_t^{(i)} \propto w_{t-1}^{(i)}\mathrm{OL}(U_t^{(i)}, \Lambda_t^{(i)})\mathrm{STP}(\Lambda_t^{(i)}; \Lambda_{t-1}^{(i)})$ and resample
4) Increment t and go back to Step 1

IV. Illumination PF-MT With Illumination Model Change

In most cases, the illumination changes gradually over time; hence, the illumination change variance takes a small value. The exception is a car or a person transitions from shadow to sunlight or vice versa or in an indoor scenario if the light bulb is switched off or on. During these transitions, if we track with a small illumination variance model, the tracker will gradually lose track. Thus, there is a need to detect model change and to assign a high illumination change variance temporarily during the transition period and to change it back once the transition is over. If we allow the illumination change to have a larger variance all the time, then the PF-MT algorithm as designed in the previous section will no longer be applicable (since it will become more likely that p^{**} is multimodal). We propose to detect model change using the recently proposed generalized (negative) ELL (gELL) statistics [7]. The gELL is designed to detect model changes before complete loss of track, which is what our application needs. In fact, it works by using the partly tracked part of the change to detect it. Standard approaches, such as tracking error, use loss of track to detect change and, hence, take longer.

gELL is the Kerridge inaccuracy [33] between the posterior at time t, $\pi_{t|t}$, and the Δ-step-ahead prediction distribution $\pi_{t|t-\Delta}$, i.e., $\mathrm{gELL}(t,\Delta) \triangleq E_{\pi_{t|t}}[-\log \pi_{t|t-\Delta}(X_t)]$ where $E_p[.]$ denotes expectation with respect to pdf $p(X)$ and $\pi_{t|t-\Delta}(X_t) \triangleq p(X_t|Y_{1:t-\Delta})$. In practical applications, it is not clear how to choose Δ. One option is to compute the maximum of gELL over all Δ, i.e., to compute $\mathrm{gELL\text{-}max}(t) \triangleq \max_{\Delta=1,2,\ldots,t}\mathrm{gELL}(t,\Delta)$. In order to detect illumination model change, we compute the gELL for the illumination state Λ_t. The gELL is computed as follows [7]. We use a Gaussian density approximation to the posterior at $t-\Delta$, i.e., $\pi_{t-\Delta|t-\Delta}(X_t)$ as $\pi_{t-\Delta|t-\Delta}^N(X_t) \approx \mathcal{N}(X_t\ \mu_{t-\Delta|t-\Delta}^N, \Sigma_{t-\Delta|t-\Delta}^N)$, where the parameters are estimated as the empirical mean and covariance of the weighted particle set for $\pi_{t-\Delta|t-\Delta}^N(X_t) = \sum_{i=1}^N w_t^{(i)}\delta(X_t - X_t^{(i)})$. With this approximation, the prediction $\pi_{t|t-\Delta}(X_t)$, which is obtained by applying the system model of Λ_t given in (4) and by multiplying Δ to $\pi_{t-\Delta|t-\Delta}(X_t)$, is also Gaussian, i.e., $\pi_{t|t-\Delta}(X_t) \approx \mathcal{N}(\mu_{t|t-\Delta}^N, \Sigma_{t|t-\Delta}^N)$ where $\mu_{t|t-\Delta}^N = \mu_{t-\Delta|t-\Delta}^N$ and $\Sigma_{t|t-\Delta}^N \triangleq \Sigma_{t-\Delta|t-\Delta}^N + \Delta\Sigma_\Lambda$. Thus,

$$\mathrm{gELL}(t,\Delta) = \sum_{i=1}^N w_t^{(i)}\left(\Lambda_t^{(i)} - \mu_{t|t-\Delta}^N\right)^T\left(\Sigma_{t|t-\Delta}^N\right)^{-1}\left(\Lambda_t^{(i)} - \mu_{t|t-\Delta}^N\right).$$

As explained in [7], the gELL threshold for detecting model change can be set at a value that is a little above $E_{\pi_{t|t-\Delta}}[-\log \pi_{t|t-\Delta}(\Lambda_t)|\text{no change}]$ (see [7, Sec. 4.3] for details). Notice that this is equal to the differential entropy of $\pi_{t|t-\Delta}(X_t)$. Since $\pi_{t|t-\Delta}(X_t)$ is approximated by a Gaussian, its differential entropy is proportional to the dimension of Λ_t times the logarithm of the determinant of the illumination change covariance.

A. Illumination PF-MT With Change Detector

We begin by running the illumination PF-MT algorithm of Algorithm 1 with Σ_Λ given by the learnt illumination covariance. At each time t, after the weighting step, we compute gELL, as described above. If it exceeds a threshold, then we set Σ_Λ to a heuristically selected large value. During this period, the tracker almost exclusively relies on the observations. Assuming no occlusion during this transition period, the particles will quickly and correctly adapt to the changed illumination conditions. At this point, the gELL statistic value will reduce. When it goes below the threshold, we reset Σ_Λ to its learnt value.

It is assumed that this transition affects the illumination space only and does not alter the observation process itself. Thus, value θ in (2) does not need to be changed. We should point out here that if occlusion occurs during this period, it will lead to tracking failure since the

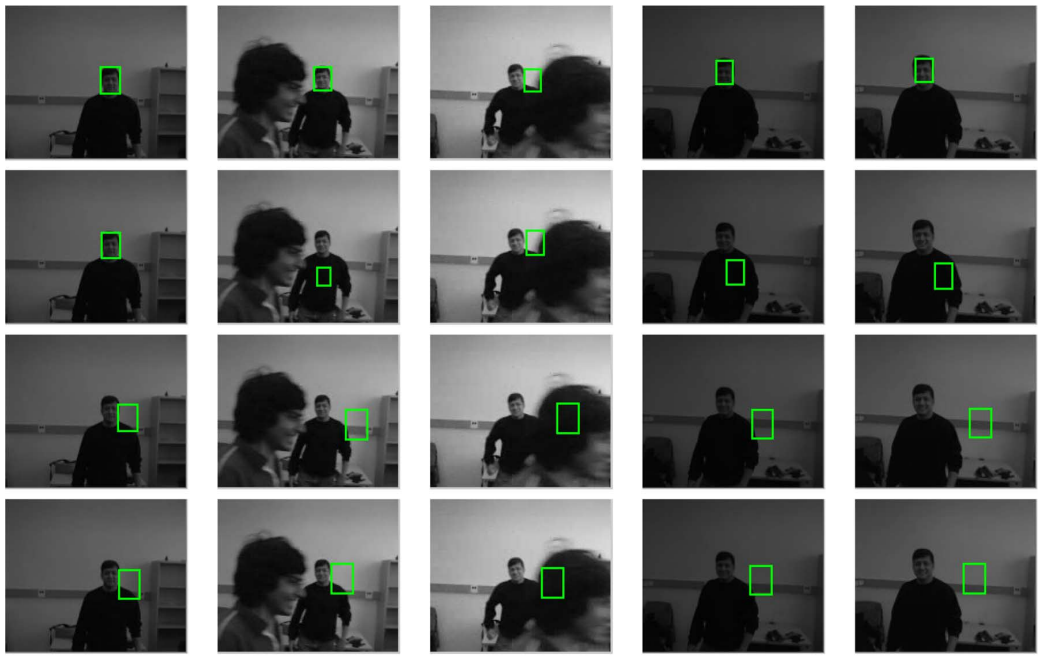

Fig. 2. Visual comparison of various methods for face tracking across illumination changes with occlusion lasting up to six frames. We used $N = 100$ particles. (top row) Illumination PF-MT (our method). (second row) Case when no model for illumination is used, i.e., PF-Gordon is used on the 3-D motion space only. (third row) Auxiliary-PF [32]. (fourth row)PF-Gordon [27] with a standard resampling strategy. It can be seen that illumination PF-MT outperforms the rest. It is to be noted that, with limited number of particles ($N = 100$), PF-Gordon looses track right from the beginning. This is because PF-Gordon fails to estimate the illumination vector correctly with insufficient number of particles.

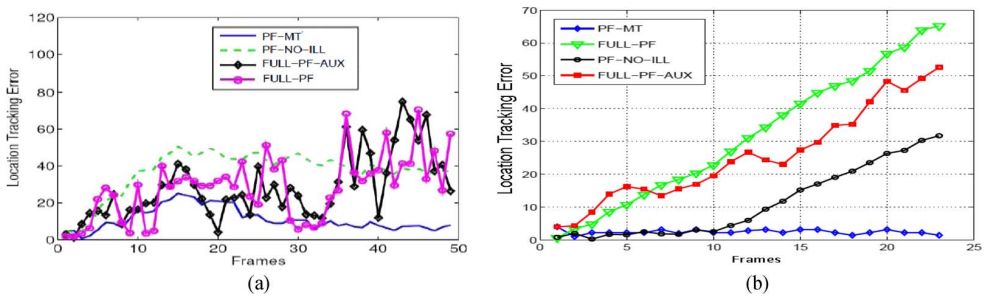

Fig. 3. Performance comparison of various PFs while tracking across illumination changes for (left) face and (right) vehicle tracking application (visuals in the supplementary section). We show the location error from the ground truth for different particle filters. PF-MT correspond to a particle filter with mode tracker (i.e., Illumination PF-MT), FULL-PF correspond to PF-Gordon [27], Full-PF-AUX correspond to Auxiliary-PF [32] and PF-NO-ILL corresponds to PF-Gordon without an illumination model. It can be seen that Illumination PF-MT outperforms the rest. It is to be noted that Auxiliary-PF has some negligible performance improvement over PF-Gordon with standard resampling strategy, but it is far worse than Illumination PF-MT. In all of these experiments, we used $N = 100$ particles. (a) Face tracking. (b) Vehicle tracking.

tracker will wrongly latch onto the occlusion intensity. In other words, the proposed solution cannot handle large illumination change and occlusion occurring at the same time.

V. Experimental Results

The goal of this correspondence is to show how to design PF-MT for illumination tracking problem and to demonstrate that it provides a much more efficient solution (efficient in terms of number of particles needed) to visual tracking under illumination change, i.e., compared with other PFs. Hence, here, we only show comparisons with other PF methods. In Fig. 3 and also in [34], we also show comparisons with some other approaches from recent work.

In all of our experiments, we used a set of labeled video sequences for learning the dynamical model parameters for Λ_t and U_t. First, we manually marked the target centroids in the training sequence and then use these to learn the motion vectors U_t. The corresponding illumination vector Λ_t is computed from the image frame Y_t as $\Lambda_t = (A^T A)^{-1} A^T Y_t(\text{ROI}(U_t))$. The covariance matrices of the change of U_t and of Λ_t, Σ_Λ, and Σ_u are estimated using the standard maximum likelihood estimation applied to $(U_t - U_{t-1})$ and $(\Lambda_t - \Lambda_{t-1})$. For learning the illumination model, we used $D = 7$ (i.e., used Legendre functions up to order 3). For all the PF algorithms, we used a fixed particle size of $N = 100$. In our experiments, we handmark the approximate target ROI in the first frame. The tracking performance of illumination PF-MT was compared with several other PF-based algorithms such as a PF without the illumination model, auxiliary-PF [32] and PF-Gordon [27]. Auxiliary-PF [32] uses look-ahead resampling strategy to improve effective particle size. PF-Doucet [9] cannot be implemented for our problem because it is not possible to use numerical convex optimization techniques to find the mode of $p^*(U_t, \Lambda_t) \propto \text{OL}(U_t, \Lambda_t)\text{STP}(U_t, \Lambda_t)$. This is due to fact that the observation likelihood $\text{OL}(U_t, \Lambda_t)$ is not continuously differentiable due to the involvement of round() operations in the mapping from U_t to Y_t [refer to (8) and (2)]. However, the same is not true for the CL of Λ_t i.e., $\text{CL}^{(i)}(\Lambda_t)$, which enables us to implement PF-MT for our problem.

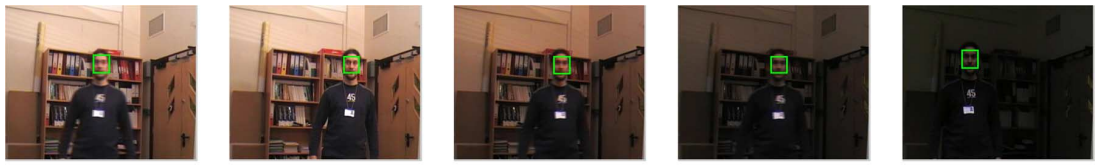

Fig. 4. Instance of face tracking under large illumination variation when someone switches the lighting conditions in a room.

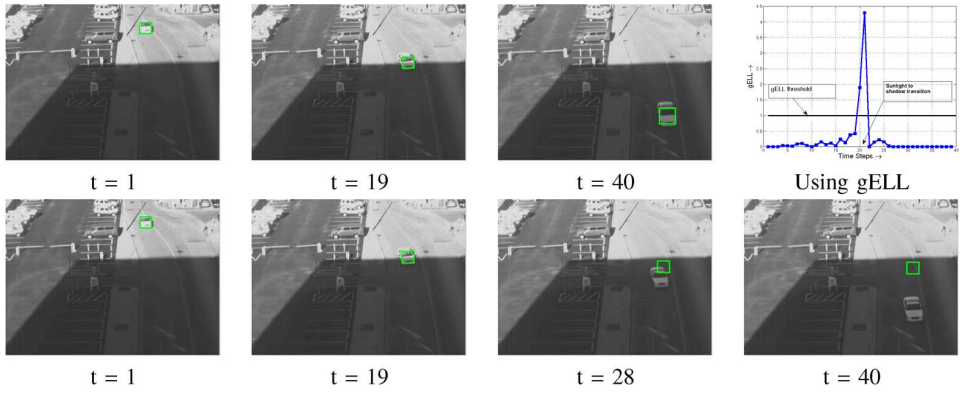

Fig. 5. Results using gELL-based change detection statistics. It can be see that the tracker can track through drastic illumination changes. (top row) Demonstrates that we are able track through illumination model changes when the car moves from sunlight to shadow area. During the transition from sunlight to shadow area (around frame 19), the gELL value goes above threshold indicating a model change (see gELL plot in the top row). (second row) The tracker fails if we do not detect the transition and keep Σ_Λ. unchanged.

In the first experiment, we evaluate the tracking performance of illumination PF-MT for face tracking in the presence of illumination change and occlusions. Here, the lighting conditions variations could be attributed to two factors: 1) the target's distance from the window and variable ambient lighting coming through it; and 2) occasional switching off and on of the light sources inside the room. The visual tracking results are given in Fig. 2. It can be clearly seen that illumination PF-MT (top row) clearly outperforms the rest of the PF-based methods for a limited particle budget of just 100 particles. The other PFs loose track within the first few frames and are unable to recover.

For quantitative tracking performance analysis, we did some further experiments with face and vehicle tracking from surveillance videos. The quantitative performance comparison plots for face tracking are shown in Fig. 3(a). The car data set was generated from a camera observing a road from above as the cars approach an intersection (shown in Fig. 1). The illumination variations were due to the variations in the ambient lighting conditions. The corresponding quantitative tracking accuracy plots are given Fig. 3(b). It can be seen that, with just 100 particles, our algorithm has the best performance in terms of tracking accuracy for both face and vehicle tracking scenarios. Another instance of face tracking under illumination variations using PF-MT has been demonstrated in Fig. 4 (data set taken from http://www.eecs.qmul.ac.uk/ andrea/avss2007_d.html). We also show visual tracking results on the standard CAVIAR data set [35] in the in Fig. 2.

We demonstrate the utility of illumination model change detection and compensation in Fig. 5. Notice that the tracker fails during a sunlight to shadow transition if we do not detect and adapt to the change.

VI. Conclusion

In this correspondence, we have tackled the difficult problem of visual tracking under variable illumination by reformulating it as a problem of large-dimensional tracking with multimodal observation likelihood and using the PF-MT approach to design an efficient PF algorithm. We show exhaustive experiments to demonstrate the superior performance of our algorithm in handling large illumination variations and severe occlusions for both face and vehicle tracking videos. We also use the recently proposed idea of gELL to detect and adapt to changes in the illumination model. In future works, sparse representation of the illumination vector could be leveraged to replace the posterior MT part by recently proposed particle filtered modified compressed sensing [36].

Appendix

The changed "appearance" of template T_t is represented in terms of a linear combination of the initial template T_0 scaled by a set of Legendre basis functions as follows [18]:

$$\begin{aligned} &\mathrm{vec}(T_t) = A\Lambda_t, \quad \text{where} \\ &\quad A \triangleq [\mathrm{vec}(T_0 \odot P_0), \ldots, \mathrm{vec}(T_0 \odot P_{D-1})] \\ &P_n(i,j) = \begin{cases} 1 & n = 0 \\ p_n(i) & n = 1, \ldots, k \\ p_{n-k}(j) & n = k+1, \ldots, D-1 \end{cases} \end{aligned} \tag{7}$$

where $p_n(.)$ is the Legendre polynomial of the nth order and Λ_t is the vector of Legendre basis coefficients at time t. Henceforth, we will call it the illumination vector. Here, A is an $M \times D$ matrix with $D = 2k+1$ with k being the highest degree of the Legendre polynomials being used, and M is the number of pixels in the initial template T_0.

Now, given the motion parameter vector U_t consisting of scale, horizontal translation, and vertical translation ($U_t = [s_t \ \tau_t^h \ \tau_t^v]^T$) of the initial template, $\mathrm{ROI}(U_t)$ can be computed as [18]

$$\begin{aligned} &\mathrm{ROI}(U_t) \triangleq \mathrm{round}\left([J_i U_t + \mathbf{i_0}, \ J_j U_t + \mathbf{j_0}]\right) \\ &\text{with} \quad J_i \triangleq \left[(\mathbf{i_0} - \tilde{i}_0 \mathbf{1}) \ \mathbf{1} \ \mathbf{0}\right], \quad J_j \triangleq \left[(\mathbf{j_0} - \tilde{j}_0 \mathbf{1}) \ \mathbf{0} \ \mathbf{1}\right]. \end{aligned} \tag{8}$$

The terms $\mathbf{i_0}$ and $\mathbf{j_0}$ are the M-dimensional vectors containing the x- and y-coordinates of all the pixels in the initial template T_0, and $\tilde{i}_0 = mean(\mathbf{i_0})$ and $\tilde{j}_0 = mean(\mathbf{j_0})$ denote the center of the initial template. Notice that (8) essentially is a geometric transformation that maps the pixels in the initial template to the current template region.

REFERENCES

[1] A. Kale, N. Vaswani, and C. Jaynes, "Particle filter with mode tracker (PF-MT) for visual tracking across illumination change," in *Proc. IEEE ICASSP*, 2007, pp. I-929–I-932.

[2] A. Kale and N. Vaswani, "Generalized ELL for detecting and tracking through illumination model changes," in *Proc. IEEE ICIP*, 2008, pp. 2736–2739.

[3] N. Vaswani, "Particle filtering for large dimensional state spaces with multimodal observation likelihoods," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4583–4597, Oct. 2008.

[4] N. Vaswani, A. Yezzi, Y. Rathi, and A. Tannenbaum, "Particle filters for infinite (or large) dimensional state spaces—Part 1," in *Proc. IEEE ICASSP*, 2006, pp. III-29–III-32.

[5] N. Vaswani, "Particle filters for infinite (or large) dimensional state spaces—Part 2," in *Proc. IEEE ICASSP*, 2006, pp. III-33–III-36.

[6] N. Vaswani, Y. Rathi, A. Yezzi, and A. Tannenbaum, "Deform PF-MT: Particle filter with mode tracker for tracking non-affine contour deformations," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 841–857, Apr. 2010.

[7] N. Vaswani, "Additive change detection in nonlinear systems with unknown change parameters," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 859–872, Mar. 2007.

[8] M. Isard and A. Blake, "Condensation: Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, Aug. 1998.

[9] A. Doucet, On sequential Monte Carlo sampling methods for Bayesian filtering Dept. Eng., Cambridge Univ., Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR.310, 1998.

[10] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[11] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of Lambertian object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1322–1333, Oct. 2002.

[12] B. P. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 1–16, 1998.

[13] G. Hager and P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 10, pp. 1025–1039, Oct. 1998.

[14] B. Han and L. Davis, "On-line density-based appearance modeling for object tracking," in *Proc. IEEE ICCV*, 2005, pp. 1492–1499.

[15] S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1491–1506, Nov. 2004.

[16] A. D. Jepson, D. J. Fleet, and T. Maraghi, "Robust online appearance models for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1296–1311, Oct. 2003.

[17] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2003, pp. 649–658.

[18] A. Kale and C. Jaynes, "A joint illumination and shape model for visual tracking," in *Proc. CVPR*, 2006, pp. 602–609.

[19] D. Comaniciu, V. Ramesh, and P. Meer, "Real time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. CVPR*, 2000, pp. 142–149.

[20] Y. Xu and A. K. Roy-Chowdhury, "Integrating motion, illumination, and structure in video sequences with applications in illumination-invariant tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 793–806, May 2007.

[21] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 125–141, May 2008.

[22] J. Jackson, A. Yezzi, and S. Soatto, "Dynamic shape and appearance modeling via moving and deforming layers," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 71–84, Aug. 2008.

[23] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1728–1740, Oct. 2008.

[24] A. Sung, T. Kanade, and D. Kim, "Pose robust face tracking by combining active appearance models and cylinder head models," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 260–274, Nov. 2008.

[25] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

[26] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. IEEE ICCV*, 2001, pp. 68–75.

[27] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng.—F (Radar Signal Process.)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[28] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[29] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[30] T. Schn, F. Gustafsson, and P. Nordlund, "Marginalized particle filters for nonlinear state-space models," *IEEE Trans. Signal Process.*, 2005.

[31] R. Chen and J. Liu, "Mixture Kalman filters," *J. R. Stat. Soc.*, vol. 62, no. 3, pp. 493–508, 2000.

[32] M. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Amer. Stat. Assoc.*, vol. 94, no. 446, pp. 590–599, Jun. 1999.

[33] D. Kerridge, "Inaccuracy and inference," *J. R. Stat. Soc., Ser. B*, vol. 23, 1961.

[34] S. Das, "Particle filtering on large dimensional state spaces and applications in computer vision" Ph.D. dissertation, Iowa State Univ., Ames, IA, 2010.

[35] The CAVIAR Dataset [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1

[36] S. Das and N. Vaswani, "Particle filtered modified compressive sensing (PaFiMoCS) for tracking signal sequences," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2010, pp. 354–358.